10-9-12 (Version II)

Genome-wide association study identifies a novel locus contributing to type 2 diabetes susceptibility in Sikhs of Punjabi origin from India

Authors:

Richa Saxena[1], Danish Saleheen[2,39,40], Latonya F Been[3], Martha L Garavito[3], Andrew Bjonnes[1], Robin Young[39], Weang Kee Ho[39], Asif Rasheed[2], Philippe Frossard[2], Xueling Sim[4,5], Neelam Hassanali [6], Venkatesan Radha[7], Manickam Chidambaram[7], Samuel Liju[7], Simon D Rees[8,9], Daniel Peng-Keat Ng[10], Tien-Yin Wong[11-13], Toshimasa Yamauchi[14], Kazuo Hara[14,15], Yasushi Tanaka[16], Hiroshi Hirose[17], Mark I McCarthy[6,18,37,38], Andrew P Morris[18,37], DIAGRAM[18], MuTHER[19], AGEN[20], Abdul Basit[21], Anthony H Barnett [9], Prasad Katulanda[6,22], David Matthews[6], Viswanathan Mohan[7,23], Gurpreet S Wander[24], Jai Rup Singh[25], Narinder K Mehra[26], Sarju Ralhan[24], M Ilyas Kamboh[27], John J Mulvihill[3], Hiroshi Maegawa[27], Kazuyuki Tobe[28], Shiro Maeda[29], Cho Y Shin[20,30], E Shyong Tai[10,31,32], M Ann Kelly[8,9], John C Chambers[33-35], Jaspal S Kooner[33,34,36], Takashi Kadowaki[14], Panos Deloukas[41], John Danesh[39], and Dharambir K Sanghera[3*]

*** Corresponding Author**

Dharambir K. Sanghera, PhD, FAHA

Department of Pediatrics, College of Medicine

University of Oklahoma Health Sciences Center

940 Stanton L. Young Blvd., Rm 317 BMSB

Oklahoma City, OK 73104, USA

Phone: 405-271-8001 ext 43064; Fax: 405-271-6027

Email: Dharambir-sanghera@ouhsc.edu

Word Count: xxxx

Number of Tables: 1 Main + 13 Supplementary

Number of Figures: 4 Main + 4 Supplementary

Affiliations

1. Center for Human Genetic Research and Department of Anesthesia, Critical Care and Pain Medicine, Massachusetts General Hospital, Harvard Medical School, Boston, MA, USA
2. Center for Non-Communicable Diseases Pakistan, Karachi, Pakistan
3. Department of Pediatrics, College of Medicine, University of Oklahoma Health Sciences Center, Oklahoma City, OK 73104, USA
4. Center for Statistical Genetics and Department of Statistics, University of Michigan, Ann Arbor, MI, USA
5. Centre for Molecular Epidemiology, National University of Singapore, Singapore
6. Oxford Centre for Diabetes, Endocrinology & Metabolism, University of Oxford, Churchill Hospital, Oxford, UK
7. Madras Diabetes Research Foundation , Chennai, India
8. College of Medical and Dental Sciences, University of Birmingham, Birmingham, United Kingdom
9. Diabetes Centre, Heart of England NHS Foundation Trust, Birmingham, United Kingdom
10. Saw Swee Hock School of Public Health, National University of Singapore, Singapore
11. Singapore Eye Research Institute, Singapore National Eye Centre, Singapore
12. Department of Ophthalmology, National University of Singapore, Singapore
13. Centre for Eye Research Australia, University of Melbourne, Melbourne, Australia
14. Department of Diabetes and Metabolic Diseases, Graduate School of Medicine, The University of Tokyo, Tokyo, Japan
15. Department of Integrated Molecular Science on Metabolic Diseases, 22nd Century Medical and Research Center, The University of Tokyo
16. Department of Internal Medicine, Division of Metabolism and Endocrinology, St. Marianna University School of Medicine
17. Health Center, Keio University School of Medicine
18. Members of the DIAGRAM study are listed in the Supplementary Online Material
19. Members of the MuTHER study are listed in the Supplementary Online Material
20. Members of the AGEN study are listed in the Supplementary Online Material
21. Baqai Institute of Diabetology and Endocrinology (BIDE), Karachi, Pakistan
22. Diabetes Research Unit, Department of Clinical Medicine, University of Colombo, Colombo, Sri Lanka
23. Dr. Mohan's Diabetes Specialities Centre, Chennai, India.
24. Hero DMC Heart Institute, Ludhiana, Punjab, India
25. Central University of Punjab, Bathinda, Punjab, India
26. All India Institute of Medical Sciences and Research, New Delhi, India
27. Department of Medicine, Shiga University of Medical Science, Singa, Japan
28. First Department of Internal Medicine, University of Toyama, Toyama, Japan
29. Laboratory for Endocrinology and Metabolism, RIKEN Center for Genomic Medicine, Kanagawa, Japan
30. Center for Genome Science, National Institute of Health, Osong Health Technology Administration Complex, Chungcheongbuk-do, Korea.
31. Department of Medicine, Yong Loo Lin School of Medicine, National University of Singapore, Singapore
32. Duke-NUS Graduate Medical School, Singapore
33. Ealing Hospital National Health Service (NHS) Trust, Middlesex, UK
34. Imperial College Healthcare NHS Trust, London, UK
35. Epidemiology and Biostatistics, Imperial College London, London, UK
36. National Heart and Lung Institute (NHLI), Imperial College London, Hammersmith Hospital, London, UK
37. Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK
38. Oxford National Institute for Health Research Biomedical Research Centre, Churchill Hospital, Oxford, UK
39. Department of Public Health and Primary Care, University of Cambridge, UK
40. Departments of Biostatistics and Epidemiology and Medicine, University of Pennsylvania, USA
41. The Wellcome Trust Sanger Institute, Hinxton, UK

ABSTRACT

We performed a genome-wide association study (GWAS) and a multi-stage meta-analysis of type 2 diabetes (T2D) in Punjabi Sikhs from India. Our discovery GWAS in 1,616 individuals (842 cases) was followed by *in silico* replication of the 513 independent SNPs ($P<10^{-3}$) in Punjabi Sikhs (n=2,819/801 cases). We further replicated 66 SNPs ($P<10^{-4}$) through genotyping in a Punjabi Sikh sample (n=2,894/1,711 cases). On combined meta-analysis in Sikh populations (n=7,329/3,354 cases), we identified a novel locus in association with T2D at 13q12 in a directly genotyped intronic SNP (rs9552911, $P= 1.82x10^{-8}$) in the *SGCG* gene. Next, we undertook *in silico* replication (Stage 2b) of the top 513 signals ($P<10^{-3}$) in 29,157 non-Sikh South Asians (10,971 cases) and *de novo* genotyping of up to 31 top signals ($P<10^{-4}$) in 10,817 South Asians (5,157 cases) (Stage 3b). In combined South Asian meta-analysis, we observed six suggestive associations ($P<10^{-5}$ to $<10^{-7}$), including SNPs at *HMG1L1/CTCFL*, *PLXNA4*, *SCAP* and chr5p11. Further evaluation of 31 top SNPs in 33,707 East Asians (16,746 cases) (Stage 3c) and 47,117 Europeans (8,130 cases) (Stage 3d), and joint meta-analysis on 128,127 individuals (44,358 cases) from 27 multi-ethnic studies did not reveal any additional loci, neither there was any evidence of replication for the new variant. Our findings provide new evidence on the presence of a population-specific signal in relation to T2D, which may provide additional insights to T2D pathogenesis.

Introduction

South Asians (people from the Indian Sub-continent) comprise more than a quarter of the global population and contribute the highest number of patients with type 2 diabetes (T2D) [1]. According to latest estimates, ~61 million people in India alone are currently affected with T2D, and their number is projected to increase to ~101 million by 2030 [2]. About 60% of the world's coronary artery disease (CAD), a principal cause of mortality in individuals with T2D, is expected to occur in India [3]. There is considerable ethnic difference in the prevalence and progression of T2D and CAD. In addition to the environmental factors, genetic factors influence disease susceptibility [4]. The incidence of T2D and CAD s about 3-5 times higher in immigrant South Asians compared to Euro-Caucasians and the age of onset of T2D is roughly a decade earlier in South Asians than in Europeans [5-7]. Higher prevalence of T2D among South Asians settled in developed countries compared to the host population reflects the genetic and ethnic predisposition to cardio-metabolic disease under an adverse environment, and joint effects of genes and environment predisposing to T2D[8]. For these reasons, we conducted ethnic-specific genetic studies in a Sikh population to dissect genetic pathways that may contribute to T2D etiology in different ethnic groups.

The vast majority of genome-wide association studies (GWAS) on T2D so far have been performed in Europeans. Studies on non-European populations, especially those with unique demographic and cultural histories are important to identify population-specific linkage disequilibrium (LD) patterns and environmental factors that may modulate disease risk or protection [9]. Interestingly, many but not all of the common loci originally identified in Europeans have been replicated in non-European groups [10-18]. Recent GWAS in non-European populations have yielded intriguing new variants [19,20] [21]. including six novel signals in South Asians represented by SNPs near *GRB14*, *ST6GAL1*, *VPS26A*, *HMG20A*, *AP3S2* and *HNF4A* in our recent meta-analysis of GWAS [22]. Given the existence of strong genetic and ethnic diversity among

South Asian communities, they do not constitute a single homogeneous community[23]. Therefore, screening South Asians with different genetic and racial background or environmental exposures may improve insights about the disease and genetic risk factors [24].

People from India have a complex racial history complicated by the presence of caste system which has prohibited inter-breeding to greater extent and has separated people into numerous endogamous groups [25]. Presently, the Sikhs, a relatively younger inbred population of ~26 million (2% of Indian population), is from the North Western province of India that follows a distinct and unique religion born 500 years ago in Punjab. They have an interesting background for 'nontraditional' disease enrichment in the absence of conventional risk factors such as smoking, obesity, and diet rich in meats [26]. Sikhs do not smoke or chew tobacco because of religious and cultural compulsions and ~50% of them are lifelong vegetarians. Despite the absence of these life style related risk factors, T2D and CAD have reached epidemic scale in Sikhs. Our initial genetic studies in a Sikh cohort as part of Asian Indian Diabetic Heart Study or the Sikh Diabetes Study (AIDHS/SDS) revealed association of *FTO* and *MTNR1B, ADIPOQ and PPARG* polymorphisms with T2D and risk factors in the absence of obesity [11,27,28]. In this investigation, we conducted a GWAS in a relatively homogenous Punjabi Sikh population of 1,850 individuals and performed multi-stage replication with up to 27 case-control studies of Punjabi, other South Asian, East Asian and Caucasian ancestries (total n=128,127; 44,358 T2D cases and 83,769 controls) **(Supplementary Tables 1,2)**. Study design of the discovery, replication, and meta-analyses phases was optimized to detect new population-specific and multi-ethnic T2D loci (**Figure 1**). One important difference in our present study from our previous South Asian GWAS [22] is that in previous study, the SNPs that were common between South Asians and Europeans were selected for replication using European sample from DIAGRAM. However, in this study, the SNPs selection was prioritized based on the top signals ($P<10^{-3}$) from our discovery Sikh cohort.

Methods

Participants

Punjabi Sikh GWAS

Study sample and characteristics

Our primary Sikh GWAS (discovery) cohort used in this investigation is comprised of 1,616 individuals from the Punjabi Sikh population which was part of the Asian Indian Diabetic Heart Study also named the Sikh Diabetes Study (AIDHS/SDS). The AIDHS/SDS has unique characteristics that are ideal for genetic studies. Sikhs are *strictly* a non-smoking population and about 50% of participants are teetotalers and life-long vegetarians. All individuals for the GWAS discovery cohort were recruited from one geographical location. Diagnosis of T2D was confirmed by scrutinizing medical records for symptoms, use of medication, and measuring fasting glucose levels following the guidelines of the American Diabetes Association [29], as described previously [11]. Lipids, insulin, glucose, anthropometric measurements, education, socio-economic status, job grade, dietary and physical activity are available on >95% of the AIDHS/SDS individuals selected for this study. Dietary questions involving alcohol consumption were scored using a scale from 0-5, details are described elsewhere[26]. T2D is often asymptomatic and remains undiagnosed for many years especially in people from the developing world due to poor healthcare provisions. Therefore, it is reasonable to assume that the actual age of onset of T2D in Sikhs may range between 39-42 years compared to the observed age at diagnosis (46 years). This age is in sharp contrast to the mean age at onset of 60 years or above in developed countries [5,26,30]. A medical record indicating either (1) a fasting plasma glucose level ≥7.0 mmol/L or ≥126 mg/dL after a minimum 12h fast or (2) a 2h post-glucose level (2h oral glucose tolerance test) ≥11.1 mmol/L or ≥200 mg/dL on more than one occasion, combined with symptoms of diabetes, confirmed the diagnosis. Impaired fasting glucose (IFG) is defined as a fasting blood glucose level ≥100 mg/dL (5.6 mmol/L) but ≤126 mg/dL (7.0 mmol/L). Impaired glucose tolerance

(IGT) is defined as a 2h OGTT > 140 mg/dL (7.8 mmol/L) but <200 mg/dL (11.1 mmol/L). Subjects with IFG or IGT were considered pre-diabetics and were excluded. The 2h OGTTs were performed following the criteria of the World Health Organizations (WHO) (75 g oral load of glucose). Body mass index (BMI) was calculated as (weight [kg]/height [meter]^2), and waist to- hip ratio (WHR) was calculated as the ratio of abdomen or waist circumference to hip circumference. Subjects with type I diabetes, or those having a family member with type I diabetes, or rare forms of T2D subtypes (maturity onset diabetes of young [MODYs]), or secondary diabetes (from e.g. hemochromatosis, pancreatitis) were excluded from the study. The selection of controls was based on fasting glycemia <100.8 mg/dL or a 2h glucose <141.0 mg/dL. Subjects with IFG or IGT were excluded when data were analyzed for association of variants with T2D. All blood samples were obtained at the baseline visits. All participants signed a written informed consent for the investigations. The study was reviewed and approved by the University of Oklahoma Health Sciences Center's Institutional Review Board, as well as the Human Subject Protection Committees at the participating hospitals and institutes in India.

South Asian cohorts

For Stage 2a replication, the Sikh component of the London Life Sciences Population (LOLIPOP) study[22] comprised 2,919 individuals (810 T2D cases and 2,018 controls) and Stage 2b the non-Sikh South Asian components of the LOLIPOP and the Pakistan Risk of Myocardial Infarction Study (PROMIS, and the Risk Assessment of Cerebrovascular Events Study [RACE]) GWA studies [22] comprised 29,157 individuals (10,971 cases and 18,186 controls) [22]. Stage 3a Punjabi-specific replication was carried out on 2,894 individuals (1,711 cases and 1,183 controls) of Punjabi ancestry from India as part of AIDHS/SDS, and replication testing amongst South Asians for Stage 3b was carried out amongst 10,817 participants (5,157 T2D cases and 5,660 controls) which were part of the following studies: Asian Indians from the Singapore Indian Eye (SINDI) study [31], the Chennai Urban Rural Epidemiology Study (CURES)[32], the Diabetes Genetics in

Pakistan (DGP) and UK Asian Diabetes Study (UKADS) [33], the Sri Lankan Diabetes Study (SLDS) [34]. Details of the contributing cohorts are provided in Supplementary Methods.

East Asian cohorts

Replication testing for Stage 3c was carried out on a total of 33,707 East Asians comprising 14,890 Japanese from RIKEN (n=7,480 genotyped) and Bio Bank Japan (n=7,410 GWAS [19,35]), and 18,817 individuals of East Asian ancestry as part of the Asian Genetic Epidemiology Network (AGEN) with genotype data available from eight GWA studies [21].

DIAGRAM (Euro-Caucasians)

Associations of SNPs with T2D amongst Europeans were tested *in silico* using results from the GWA phase of the DIAbetes Genetics Replication And Meta-analysis (DIAGRAM) study comprising 47,117 subjects[36].

Genotyping and quality control

Genomic DNA was extracted from buffy coats using QiaAmp blood kits (Qiagen, Chatworth, USA) or by the salting out procedure[37]. Stage 1 genome-wide genotyping was performed using a Human 660W Quad BeadChip panel (Illumina, USA). We performed pair wise identity-by-state (IBS) clustering in PLINK across all individuals to assess population stratification; no population outliers were detected. Related individuals with pi-hat > 0.3 and samples with <95% call rate were excluded as were SNPs with call rate <95%, also excluded were SNPs with Hardy-Weinberg equilibrium (HWE) $P<10^{-6}$ or minor allele frequency <1%. After quality control, 524,216 directly genotyped SNPs in 1,616 subjects (842 cases and 774 controls) were available for association testing. Genotyping for *de novo* SNPs in the replication samples was performed by Sequenom MassArray, BioMark HD MX/HX Genetic Analysis System by Fluidigm, or KASPAR (K-Bioscience Ltd, UK). Samples and SNPs with <95% call rate were excluded, or

that deviated from HWE at $P<10^{-3}$. The associations of SNPs with T2D were tested in each cohort separately.

Statistical analyses

Association testing

Associations of SNPs with T2D were tested using logistic regression and an additive genetic model. Gender, age, BMI and five, or ten principal components to adjust for residual population stratification were included as covariates. As the existing HapMap2 or HapMap3 and 1000 Genomes data do not include Sikhs, the five or ten principal components used for this correction were estimated using our Sikh population sample and not from HapMap populations. After association analyses the genomic control inflation factor, lambda was 1.0, so no adjustments were made **(Supplementary Figures 2 a, b).**

In addition to analysis of directly genotyped SNPs, we performed imputation using Impute 2 program[38-40], which determines the probability distribution of missing genotypes based on a set of known haplotypes and an estimated fine-scale recombination map. Imputation was based on the entire multi-ethnic HapMap3 reference panel of ~1.5 million autosomal SNPs with MAF>0.01 in 1,011 individuals from Africa, Asia, Europe and the Americas (including 1,362,138 SNPs from the Indian population of 100 Gujaratis from Houston [GIH]). Imputation yielded a total of 1,241,769 total passing SNPs with MAF > 1% in the Sikh GWAS. Imputed SNPs were analyzed using SNPTEST [38,40], adjusted for covariates age, gender, BMI and five principal components, which implements frequentist tests that calculate *P* values, parameter estimates and their standard errors that account for the uncertainty due to the probability distributions of the imputed genotypes, and included only those SNPs with an information score of >= 0.5 in discovery sample as well as in all GWAS studies used for replication, a measure of the relative statistical information about the additive genetic effect being estimated. The genomic control value for imputed SNPs was 1.02.

The inbreeding coefficient and measures of autozygosity were determined using the program PLINK. We identified runs of homozygosity using the metrics defined in Nalls et al.[41], evaluating 1 Mb autosomal regions with at least 50 adjacent SNPs, with a sliding window of 50 SNPs including no more than 2 SNPs with missing genotypes and 1 possible heterozygous genotype.

Stage 2 Replication: We selected all independent association signals (r^2<0.25) with $P<10^{-3}$ for lookup in GWAS of a) the Sikh component of the LOLIPOP GWAS study [22] and b) the non-Sikh South Asian components of the LOLIPOP and PROMIS GWAS studies [22]. A fixed-effect, inverse variance meta-analysis (as implemented in METAL) [42] was used to combine the results for individual studies.

Stage 3 Replication: Significant association results with $P<10^{-3}$ based on meta-analysis of stages 1, 2a and 2b were selected for *de novo* or *in silico* replication in Sikh, South Asian, other Asian, and European populations. In addition, we selected SNPs from a Sikh-only meta-analysis of stages 1a and 2a for genotyping in an in-house Punjabi Sikh T2D case-control population. Signals with $P<10^{-4}$ after meta-analysis of stages 1, 2a and 3a were also genotyped in the South Asian, other Asian and European populations to test if they were specific to the Sikh ethnic group or spanned ethnicities. All meta-analyses were performed using a fixed-effects inverse variance meta-analysis implemented in METAL.

MuTHER Consortium

The MuTHER resource (www.muther.ac.uk) includes LCLs, skin and adipose tissue derived simultaneously from a subset of well-phenotyped healthy female twins from the Twins UK adult registry. Whole-genome expression profiling of the samples, each with either two or three technical replicates, were performed using the Illumina Human HT-12 V3 BeadChips (Illumina Inc) according to the protocol supplied by the

manufacturer. Log2 transformed expression signals were normalized separately per tissue as follows: quantile normalization was performed across technical replicates of each individual followed by quantile normalization across all individuals. Genotyping was done with a combination of Illumina arrays HumanHap300, HumanHap610Q, 1M-Duo and 1.2MDuo 1M. Untyped HapMap2 SNPs were imputed using the IMPUTE software package (v2). The number of adipose samples with genotypes and expression values is 776. Association between all SNPs (MAF>5%, IMPUTE info >0.8) within a gene or within 1MB of the gene transcription start or end site and normalized expression values were performed with the GenABEL/ProbABEL packages using the polygenic linear model incorporating a kinship matrix in GenABEL followed by the ProbABEL mmscore score test with imputed genotypes. Age and experimental batch were included as cofactors.

Results

Punjabi Sikh discovery GWAS

Clinical characteristics of the Stage 1 Punjabi Sikh T2D GWAS cohort and Stage 2a and 2b (replication) cohorts are described in **Supplementary Table 3**. Principal components analysis revealed little population structure **(Supplementary Figure 1).** After quality control, 524,216 directly genotyped SNPs in 1,616 subjects (842 cases and 774 controls) from 1,850 total subjects were available for association testing after removing samples showing cryptic relatedness through identity by descent sharing. To increase genome coverage, genotypes were imputed for un-typed SNPs using the HapMap3 multi-ethnic reference panel **(see Methods)**, yielding a total of 1,241,769 SNPs for association analyses. The reason for choosing a more cosmopolitan panel and not restricting to the GIH was based on our own data showing equal diversity of the Sikhs from GIH and CEU, and based on previously described advantages of using a worldwide reference panel [39]. We performed GWAS for T2D adjusted for covariates age, gender, BMI and

five principal components **(Supplementary Figure 1)**; no evidence of inflation was observed **(Supplementary Figure 2 a,b) (see Methods) .**

Replication and meta-analyses in Punjabi Sikh participants

We undertook a two-stage replication in T2D case-control samples of Punjabi Sikh ancestry (Stages 2a & 3a in **Figure 1**). Lead SNPs representing 513 novel independent (r^2<0.25) association signals with P <10^{-3} in the discovery GWAS (including only two previously known GWAS SNPs from *TCF7L2* and *IGF2BP2* and excluding 62 SNPs with P<10^{-3} from other known T2D loci) were tested for *in silico* replication in the Punjabi Sikh sub-component of the London Life Sciences Population (LOLIPOP) GWAS study comprising 801 T2D cases and 2,018 controls **(Supplementary Table 1)**. Top SNPs representing 66 putatively novel signals with P<10^{-4} after Stage 1 & 2a meta-analysis using a fixed effects, inverse-variance approach were directly genotyped in the stage 3a sample of 2,894 Punjabi Sikh individuals (1,711 T2D cases and 1,183 controls) **(Figure 1, Supplementary Table 2).**

In a combined meta-analysis of the three Punjabi studies (n=7,329), we identified one new locus reaching genome-wide significance (P< 5x10^{-8}) along with robust replication of the established SNP rs7903146 in *TCF7L2 (P*= 3.32x10^{-19}) in Sikhs **(Figures 2-4)**. This novel association signal lies in a 164 kb region of strong LD at 13q12 (harboring genes sarcoglycan, gamma *SGCG* and sacsin, *SACS)* and is represented by a directly genotyped intronic SNP, rs9552911 in *SGCG* (OR=0.67, 95%CI [0.58-0.77], P= 1.82x10^{-8}) for the minor 'A' allele **(Table 1, Figure 4, Supplementary Table 5).** Excluding BMI from logistic regression model did not affect the association **(Supplementary Table 6).** Furthermore, including 5 additional principal components in the model did not attenuate the signal, indeed, the effect and significance was slightly improved (**Supplementary Table 6)**. The genetic variance (R^2) explained by this variant for T2D phenotype in Punjabi Sikh discovery and replication sets was 1.57%, 1.34%, respectively. There were 15 additional independent

loci with suggestive evidence ($P<10^{-5}$ to $<10^{-7}$) of association included *six* unknown regions along with *IGF2BP2* originally identified in Caucasians[43] **(Supplementary Table 5).** Meta-analysis results including non-Sikh Punjabis from PROMIS (Pakistan) revealed suggestive association ($P<10^{-5}$ to $<10^{-7}$) at SNPs from three new regions: chromosome 18q21 *ZBTB7C* (rs1893835), 20q13, near *HMG1L1/CTCFL/RBM38/PCK1* (rs328506), and at 5q33 (rs17053082) **(Supplementary Table 7)**.

Replication/evaluation and meta-analysis in other South Asians

In order to identify T2D association signals common to Punjabi and other South Asian populations, we tested for association of the 513 top independent signals ($P<10^{-3}$) derived from the discovery cohort in GWAS from the LOLIPOP, PROMIS, and RACE studies as part of Stage 2b replication (10,971 T2D cases and 18,186 controls) **(Figure 1, Supplementary Table 1**). 31 signals ($P<10^{-4}$ from an interim analysis with Stage 2b) were further genotyped in 10,817 South Asians (5,157 T2D and 5,660 controls) **(Figure 1)** as part of Stage 3b replication. Clinical characteristics of the Stage 3 replication cohorts are described in **Supplementary Table 4**. Combined South Asian meta-analysis revealed nominally significant association in 6 SNPs with MAF >5% ($P\leq10^{-4}$), but only the two previously known SNPs in *TCF7L2* and *IGF2BP2* reached genome-wide significance **(Table 1, Supplementary Table 8)**. These suggestive novel signals included SNPs at chromosome 20q13, near *HMG1L1/CTCFL/RBM38/PCK1* (rs328506), 7q32 near *PLXNA4* (rs1593304), 3p21 in *SCAP* (rs4858889) and at 5p11 represented by rs13155082 **(Supplementary Table 8)**. Further studies and replication in a larger sample will be required to validate these results and identify causal variants at these loci.

Multi-ethnic replication and meta-analysis

To identify T2D signals spanning ethnicities, we extended replication of 31 SNPs with $P<10^{-4}$ in Punjabis and South Asians (Stage 3b) to East Asians (AGEN+) and Europeans (DIAGRAM+) in Stages 3c and 3d

respectively **(Figure 1)**. Upon meta-analysis of 31 loci in Asians (South Asians and AGEN+), genome-wide associations were only seen in *TCF7L2* (rs7903146, P=1.93 x10^{-38}) and *IGF2BP2* (rs1470579, P=1.54x10^{-13}) **(Supplementary Table 9)**. In joint multi-ethnic meta-analysis on 128,127 individuals from 27 studies, only two previously known loci *TCF7L2* (rs7903146, P=8.53x10^{-75}) and *IGF2BP2* (rs1470579, P=1.81x10^{-19}) showed robust associations. Interestingly, none of the Punjabi hits could be independently confirmed in AGEN+ or DIAGRAM+ (notably the lead rs9552911 variant from *SGCG* was monomorphic in DIAGRAM+) **(Table 1, Supplementary Table 10)**. Lookup of 50 Kb upstream and downstream of SNPs within the *SGCG* locus in the publicly available data of MAGIC Study on glycemic trait GWAS [44,45] revealed several nominal associations of SNPs with fasting blood glucose and two hour glucose levels **(Supplementary Figure 3)**. Some of these SNPs also showed association with fasting blood glucose and waist or WHR in Sikhs **(Supplementary Table 11)** , but none of these were in LD (r^2>0.20) with our lead SNP**.**

Gene Expression studies

We examined expression of *SGCG* and neighboring genes *(FLJ46358, MIPEP, SACS, TNFRSF19)* within 1 Mb of the index SNP by cis-expression quantitative trait locus (eQTL) analysis using adipose tissue, skin and lymphoblastic cell lines (LCL) gene expression data from the MuTHER consortium comprising healthy female twins of European ancestry from Britain. (www.muther.ac.uk). Several SNPs in the *SGCG* region were associated with significantly elevated ($P_{eQTL}$$10^{-4}$ to 10^{-9}) expression of SGCG mRNA in adipose tissues **(Supplementary Table 12, Supplementary Figure 4).** One adipose eQTL from MuTHER (rs572303; $P_{eQTL=}$5.47x10^{-4}) located within *SGCG* showed a nominally significant association with increased waist circumference in Sikhs (β=0.67, P= 5.2x10^{-2}) (Supplementary Table 11). As shown in **Supplementary Figure 4,** the LD patterns in the region (~1.46 Mb) surrounding *SGCG* variant (rs9552911) varied in East Asians (JPT), Africans (YRI), Caucasians (CEU), Gujarati Indians (GIH) and Sikhs. Interestingly, in Caucasians and Yorubians, this variant was monomorphic. However, several alternative SNPs from this region in

Europeans were associated with fasting blood glucose (MAGIC study, r^2 ranging from 0.10-0.20) and mRNA expression in adipose tissues in MuTHER Study (r^2 ranging from 0.14-0.26) with the index SNP (rs9552911). These data suggest that the population differences may underlie the weak LD. It is possible that a single causal variant may be responsible for these associations but LD may differ between Sikhs, Europeans, and other populations.

Comparative analysis of autozygosity

We further to compare distributions of inbreeding coefficients and autozygosity as described by Nalls et al,[41]. As expected, the inbreeding coefficients in our sample were higher compared to two outbred populations of European Americans- Coriell and BLSA (F=0.041 +/- 0.018 in Sikhs vs. F=0.007 +/- 0.019 in Coriell and F =-0.3 +/- 0.012 in BLSA), as assessed by Nalls et al,[41]. However, these results were similar to other Indian populations previously reported by Reich et al.[46]. No significant difference in inbreeding was observed between cases and controls (*P*=0.59). Autozygosity analysis determined that there were 19 +/- 7 homozygous segments over 1 Mb in length, with average length of 2.0 +/- 0.95 Mb. Hence, fewer but longer autozygous segments were found than in outbred populations. No correlation of measures of autozygosity to age was observed (*P*>0.05) across decades of age.

Discussion and conclusions

In this GWAS and a multi-stage meta-analysis, a novel locus at 13q12 in the *SGCG* gene (rs9552911; *P*= $1.82x10^{-8}$) was identified to be associated with T2D susceptibility in Punjabi Sikhs from Northern India. *SGCG* is a member of sarcoglycan complex of transmembrane glycoproteins associated with autosomal recessive muscular dystrophy, in particular limb-girdle muscular dystrophy type 2C (LGMD2C). *SGCG* is expressed in skeletal muscles and its high expression is also seen in vascular smooth muscle cells as well as in breast cancer cell lines [47,48]. Founder mutations in *SGCG* causing LGMD2C predate migration of

Romani gypsies of Europe out of India around 1100 AD [49]. Due to complete endogamy, this genetically isolated community had an increased incidence of autosomal recessive LGMD2C. SGCG targeted knockout mice displayed a variety of phenotypes including dystrophic cardiomyopathy, defects in skeletal muscle, metabolism, homeostasis, growth, apoptosis, aging, and behavior [50-53]. Mice lacking sarcoglycan complex including the *SGCG* in adipose and skeletal muscle were shown to be glucose intolerant and exhibited whole body insulin resistance due to impaired insulin-stimulated glucose uptake in skeletal muscles [54].

The allelic distribution of the less common 'A' (protective) allele of rs9552911 ranged from 0.06 to 0.15 in South Asians and differed among other South Asians (0.11) and Punjabi Sikhs (0.08) for this SNP **(see details in Supplementary Table 13).** Perhaps allele frequency difference may account for the lack of replication in non-Sikh South Asians. Further replication in large independent datasets of South Asians and Punjabi Sikhs would be needed to explain this observed association. In view of the complex racial history complicated by a well-defined caste system, Indian populations display a great deal of genetic and cultural diversity [55]. Studies suggest that genetic affinity among endogamous communities in India is inversely correlated with geographic distance between them [23]. Therefore, it is possible that undetected causal variant(s) or multiple rare variants in LD with this marker arose on a haplotype tagged by rs9552911 in Punjabi Sikhs after divergence from other South and West Indian populations. This variation in the index SNP rs9552911 does not appear to be of recent origin as suggested by comparative genomic analysis **(Supplementary Figure 5).** Two important nuclear hormone receptors and transcription factors (PPAR gamma (1 and 2) and PPAR alpha) bind on the promoter and intron 1 of the *SGCG*. Further, the maturity onset diabetes of young-4 [MODY4] locus at chromosome 13q12 represented by insulin promoter factor 1 or *PDX-1* lies next to the *SGCG* locus. Therefore, further in-depth examination by targeted resequencing in the extended region and functional studies may reveal putative causative variants in this extended region and provide insight into the physiological relevance of the observed association.

In summary, our study identified a novel locus associated with T2D in a population of Punjabi Sikh ancestry from Northern India. These findings provide new information not only on previously unknown regions associated with T2D but demonstrate a putative population-specific association which could lead to additional biological insights into T2D pathogenesis.

Acknowledgments

This work was supported by the National Institutes of Health grant NIDDK R01DK082766, funded by National Institute of Diabetes and Digestive and Kidney Diseases and a seed grant from University of Oklahoma Health Sciences Center, Oklahoma City, USA. We thank the research participants for their contribution and support for making this study possible. We also thank Bansari Mehta, Braden Juengel, and Tim Braun from Oklahoma University Health Sciences Center and Jackie Lane from Harvard Medical School for their superb technical support in this study. We thank the collaborating groups of all the contributing sites including Dr. Nadereh Jafari and staff at the Northwestern University's Genomic Core facility for carrying out genome-wide genotyping and Dr. Andrew Brooks and staff at the Rutgers University for replication genotyping**.** Dr. Sanghera is the guarantor of this work, had full access to all the data, and takes full responsibility for the integrity of data and the accuracy of data analysis. **Study-specific acknowledgments are provided in the Supplementary Online Material.**

Author Contributions

Manuscript preparation: R.S. and L.F.B performed analysis. D.K.S., R.S., and D.S. wrote the manuscript. All authors read and provided critical comment on the manuscript. Data collection and analysis in the participating studies: AGEN Consortium: Y.S.C. and M. S., Asian Indian Diabetic Heart study/Sikh Diabetes Study Sikh Diabetes Study: L.F.B., M.L.G., D.K.S, R.S., M.I.K., J.J.M., G.S.W, J.R.S., N.K.M, and S.R., Bio Bank Japan and RIKEN T2D Study: T.Y., K.H., Y.T., H.H., H.M., K.T., S.M., and T.K., Chennai Urban Rural

Epidemiology Study: V.R., M. C, S.L. and V.M., Diabetes Genetics in Pakistan and UK Asian Diabetes Studies: S.D.R., A.B., A.H.B. and M.A.K., London Life Sciences Population Study: J.S.K., and J.C.C., Pakistan Risk of Myocardial Infarction Study and Risk Assessment of Cerebrovascular Events Study: D.S., P. F, A.R., R.Y., M.W. J.D. and D.J.R., M, Singapore Indian Eye Study: X.S., D.P.N., T.Y.W., and E.S.T., Sri Lankan Diabetes Study: N.H., D.M., P.K. and M.I.M. Association results among Europeans in DIAGRAM: A.P.M. and M.I.M. eQTL analyses in MuTHER: A.S.D., K.S.S. and M.I.M.

Declaration of Interest

We declare that there is no conflict of interest that could be perceived as prejudicing the impartiality of the research reported.

URLs

IMPUTE http://mathgen.stats.ox.ac.uk/impute/impute.html;
METALhttp://www.sph.umich.edu/csg/abecasis/Metal; SNAP http://www.broadinstitute.org/mpg/snap/;
LocusZoom, http://csg.sph.umich.
edu/locuszoom/; GenABEL, http://www.genabel.org/; ProbABEL,
http://www.genabel.org/packages/ProbABEL.

References

1. **Basnyat, B. & Rajapaksa, L.C. Cardiovascular and infectious diseases in South Asia: the double whammy. *British Medical Journal* 328, 781 (2004).**
2. **N, U., D, W., L, G., D, G. & (Editors) (eds.). *Diabetes Atlas*, (International Diabetes Federation, Belgium, 2011).**
3. **Ghaffar, A., Reddy, K.S. & Singhi, M. Burden of non-communicable diseases in South Asia. *BMJ* 328, 807-10 (2004).**
4. **Zimmet, P. Type 2 (non-insulin-dependent) diabetes--an epidemiological overview. *Diabetologia* 22, 399-411 (1982).**
5. **McKeigue, P.M., Pierpoint, T., Ferrie, J.E. & Marmot, M.G. Relationship of glucose intolerance and hyperinsulinaemia to body fat pattern in south Asians and Europeans. *Diabetologia* 35, 785-91 (1992).**
6. **Oldroyd, J., Banerjee, M., Heald, A. & Cruickshank, K. Diabetes and ethnic minorities. *Postgrad Med J* 81, 486-90 (2005).**
7. **Nakagami, T. et al. Age, body mass index and Type 2 diabetes-associations modified by ethnicity. *Diabetologia* 46, 1063-70 (2003).**
8. **Zimmet, P.Z. Kelly West Lecture 1991. Challenges in diabetes epidemiology--from West to the rest. *Diabetes Care* 15, 232-52 (1992).**

9. McCarthy, M.I. et al. Genome-wide association studies for complex traits: consensus, uncertainty and challenges. *Nat Rev Genet* 9, 356-69 (2008).
10. Sanghera, D.K. et al. TCF7L2 polymorphisms are associated with type 2 diabetes in Khatri Sikhs from North India: genetic variation affects lipid levels. *Ann Hum Genet* 72, 499-509 (2008).
11. Sanghera, D.K. et al. Impact of nine common type 2 diabetes risk polymorphisms in Asian Indian Sikhs: PPARG2 (Pro12Ala), IGF2BP2, TCF7L2 and FTO variants confer a significant risk. *BMC Med Genet* 9, 59 (2008).
12. Hu, C. et al. A genetic variant of G6PC2 is associated with type 2 diabetes and fasting plasma glucose level in the Chinese population. *Diabetologia* 52, 451-6 (2009).
13. Hu, C. et al. PPARG, KCNJ11, CDKAL1, CDKN2A-CDKN2B, IDE-KIF11-HHEX, IGF2BP2 and SLC30A8 are associated with type 2 diabetes in a Chinese population. *PLoS One* 4, e7643 (2009).
14. Ng, M.C. et al. Implication of genetic variants near TCF7L2, SLC30A8, HHEX, CDKAL1, CDKN2A/B, IGF2BP2, and FTO in type 2 diabetes and obesity in 6,719 Asians. *Diabetes* 57, 2226-33 (2008).
15. Sanghera, D.K. et al. Testing the association of novel meta-analysis-derived diabetes risk genes with type II diabetes and related metabolic traits in Asian Indian Sikhs. *J Hum Genet* 54, 162-8 (2009).
16. Omori, S. et al. Replication study for the association of new meta-analysis-derived risk loci with susceptibility to type 2 diabetes in 6,244 Japanese individuals. *Diabetologia* 52, 1554-60 (2009).
17. Schleinitz, D. et al. Lack of significant effects of the type 2 diabetes susceptibility loci JAZF1, CDC123/CAMK1D, NOTCH2, ADAMTS9, THADA, and TSPAN8/LGR5 on diabetes and quantitative metabolic traits. *Horm Metab Res* 42, 14-22.
18. Rees, S.D. et al. Replication of 13 genome-wide association (GWA)-validated risk variants for type 2 diabetes in Pakistani populations. *Diabetologia* 54, 1368-74.
19. Yamauchi, T. et al. A genome-wide association study in the Japanese population identifies susceptibility loci for type 2 diabetes at UBE2E2 and C2CD4A-C2CD4B. *Nat Genet* 42, 864-8.
20. Unoki, H. et al. SNPs in KCNQ1 are associated with susceptibility to type 2 diabetes in East Asian and European populations. *Nat Genet* 40, 1098-102 (2008).
21. Cho, Y.S. et al. Meta-analysis of genome-wide association studies identifies eight new loci for type 2 diabetes in east Asians. *Nat Genet* 44, 67-72.
22. Kooner, J.S. et al. Genome-wide association study in individuals of South Asian ancestry identifies six new type 2 diabetes susceptibility loci. *Nat Genet* 43, 984-9.
23. Kivisild, T. et al. Deep common ancestry of indian and western-Eurasian mitochondrial DNA lineages. *Curr Biol* 9, 1331-4 (1999).
24. Mathew, C.G. New links to the pathogenesis of Crohn disease provided by genome-wide association scans. *Nat Rev Genet* 9, 9-14 (2008).
25. Singh, K., Bhalla, V. & Kaul, L. *People of India.*, (Oxford University Press, Vol X :p1-3. , 1994).
26. Sanghera, D.K. et al. The Khatri Sikh Diabetes Study (SDS): study design, methodology, sample collection, and initial results. *Hum Biol* 78, 43-63 (2006).
27. Sanghera, D.K. et al. PPARG and ADIPOQ gene polymorphisms increase type 2 diabetes mellitus risk in Asian Indian Sikhs: Pro12Ala still remains as the strongest predictor. *Metabolism* 59, 492-501.
28. Been, L.F. et al. A low frequency variant within the GWAS locus of MTNR1B affects fasting glucose concentrations: Genetic risk is modulated by obesity. *Nutr Metab Cardiovasc Dis*.

29. **Diagnosis and classification of diabetes mellitus.** ***Diabetes Care*** **27 Suppl 1, S5-S10 (2004).**
30. **McKeigue, P.M. et al. Diabetes, hyperinsulinaemia, and coronary risk factors in Bangladeshis in east London.** ***Br Heart J*** **60, 390-6 (1988).**
31. **Lavanya, R. et al. Methodology of the Singapore Indian Chinese Cohort (SICC) eye study: quantifying ethnic variations in the epidemiology of eye diseases in Asians.** ***Ophthalmic Epidemiol*** **16, 325-36 (2009).**
32. **Chidambaram, M., Radha, V. & Mohan, V. Replication of recently described type 2 diabetes gene variants in a South Indian population.** ***Metabolism*** **59, 1760-6.**
33. **Rees, S.D. et al. An FTO variant is associated with Type 2 diabetes in South Asian populations after accounting for body mass index and waist circumference.** ***Diabet Med*** **28, 673-80.**
34. **Katulanda, P. et al. Prevalence and projections of diabetes and pre-diabetes in adults in Sri Lanka--Sri Lanka Diabetes, Cardiovascular Study (SLDCS).** ***Diabet Med*** **25, 1062-9 (2008).**
35. **Gille, H., Strahl, T. & Shaw, P.E. Activation of ternary complex factor Elk-1 by stress-activated protein kinases.** ***Curr Biol*** **5, 1191-200 (1995).**
36. **Voight, B.F. et al. Twelve type 2 diabetes susceptibility loci identified through large-scale association analysis.** ***Nat Genet*** **42, 579-89.**
37. **Miller, S.A., Dykes, D.D. & Polesky, H.F. A simple salting out procedure for extracting DNA from human nucleated cells.** ***Nucleic Acids Res*** **16, 1215 (1988).**
38. **Marchini, J., Howie, B., Myers, S., McVean, G. & Donnelly, P. A new multipoint method for genome-wide association studies by imputation of genotypes.** ***Nat Genet*** **39, 906-13 (2007).**
39. **Howie, B., Marchini, J. & Stephens, M. Genotype imputation with thousands of genomes.** ***Genes, Genomics, Genetics*** **1, 457-70 (2011).**
40. **Marchini, J. & Howie, B. Genotype imputation for genome-wide association studies.** ***Nat Rev Genet*** **11, 499-511.**
41. **Nalls, M.A. et al. Measures of autozygosity in decline: globalization, urbanization, and its implications for medical genetics.** ***PLoS Genet*** **5, e1000415 (2009).**
42. **Willer, C.J., Li, Y. & Abecasis, G.R. METAL: fast and efficient meta-analysis of genomewide association scans.** ***Bioinformatics*** **26, 2190-1.**
43. **Sladek, R. et al. A genome-wide association study identifies novel risk loci for type 2 diabetes.** ***Nature*** **445, 881-5 (2007).**
44. **Dupuis, J. et al. New genetic loci implicated in fasting glucose homeostasis and their impact on type 2 diabetes risk.** ***Nat Genet*** **42, 105-16.**
45. **Saxena, R. et al. Genetic variation in GIPR influences the glucose and insulin responses to an oral glucose challenge.** ***Nat Genet*** **42, 142-8.**
46. **Reich, D., Thangaraj, K., Patterson, N., Price, A.L. & Singh, L. Reconstructing Indian population history.** ***Nature*** **461, 489-94 (2009).**
47. **Avondo, F. et al. Fibroblasts from patients with Diamond-Blackfan anaemia show abnormal expression of genes involved in protein synthesis, amino acid metabolism and cancer.** ***BMC Genomics*** **10, 442 (2009).**
48. **Barresi, R., Moore, S.A., Stolle, C.A., Mendell, J.R. & Campbell, K.P. Expression of gamma -sarcoglycan in smooth muscle and its interaction with the smooth muscle sarcoglycan-sarcospan complex.** ***J Biol Chem*** **275, 38554-60 (2000).**
49. **Piccolo, F. et al. A founder mutation in the gamma-sarcoglycan gene of gypsies possibly predating their migration out of India.** ***Hum Mol Genet*** **5, 2019-22 (1996).**
50. **Goldstein, J.A. et al. SMAD signaling drives heart and muscle dysfunction in a Drosophila model of muscular dystrophy.** ***Hum Mol Genet*** **20, 894-904.**

51. Townsend, D., Yasuda, S., McNally, E. & Metzger, J.M. Distinct pathophysiological mechanisms of cardiomyopathy in hearts lacking dystrophin or the sarcoglycan complex. *FASEB J* 25, 3106-14.
52. Wheeler, M.T. et al. Secondary coronary artery vasospasm promotes cardiomyopathy progression. *Am J Pathol* 164, 1063-71 (2004).
53. Hack, A.A. et al. Gamma-sarcoglycan deficiency leads to muscle membrane defects and apoptosis independent of dystrophin. *J Cell Biol* 142, 1279-87 (1998).
54. Groh, S. et al. Sarcoglycan complex: implications for metabolic defects in muscular dystrophies. *J Biol Chem* 284, 19178-82 (2009).
55. Bamshad, M. et al. Genetic evidence on the origins of Indian caste populations. *Genome Res* 11, 994-1004 (2001).

Figure Legends

Figure 1. Summary of study design and outcome of key findings

Figure 2a and 2b. Manhattan plot in Figure 2a show primary genome-wide association analysis of the Punjabi Sikh discovery cohort using directly genotyped (533,977) SNPs and Manhattan plots in Figure 2b show imputed 1,241,769 SNPs on X axis and -log10 p value of association on Y axis. Locations of the three loci (including one novel locus at the *SGCG*) reached genome-wide significance after combined analysis of the GWAS and replication data in Punjabi Sikhs.

Figures 3a and 3b. Figure 3a shows regional association plot for a new T2D locus detected at 13q12 in the SGCG gene in from the genome-wide meta-analysis in Sikhs. Figure 3b shows a strong confirmation of SNPs in the TCF7L2 in Sikh meta-analysis. In these plots, for SNP showing the most strongly associated signal are depicted as a red diamond for the combined stage 1, 2a and 3a results for meta-analysis and purple diamond shows evidence of association for the stage 1 results. Each circle in color shows a SNP with a color- scale relating the r^2 value for that SNP and the top SNP taken from HapMap 3 GIH panel. We present LD using the GIH panel, the closest HapMap population to the Sikhs; however, we note that there could still be differential LD between the reference panel and the Sikh population. At the bottom of the plot, the locations of known genes in the region are shown.

Figure 4: Forest plot showing association of lead SNP in the *SGCG* (rs9552911) with type 2 diabetes. For each study, the estimates of odds ratios with 95% CI are shown. In addition, meta-analysis of Sikhs, South Asians, and multi-ethnic studies are shown at the bottom. Meta-analysis in Sikhs show a significant association of rs9552911 with T2D (OR 0.67[0.58-0.77], P=1.82x10^{-8}]

Table 1. GWAS, replication, and meta-analysis results of T2D loci identified in Punjabi Sikhs

SNP	Chr. Position	Nearest Gene	Effect/ Other Allele	EAF[a] Sikh[b]	EAF[a] CEU[c]	Punjabi Sikh Discovery GWAS (Stage 1) OR (95%CI) P value	Punjabi Sikh Replication (Stages 2a, 3a) OR (95%CI) P value	Other South Asian Replication (Stages 2b, 3b) OR (95%CI) P value	East Asian Evaluation (Stage 3c) OR (95%CI) P value	European Evaluation (Stage 3d) OR (95%CI) P value	Punjabi meta-analysis OR (95%CI) P value	South Asian meta-analysis OR (95%CI) P value	Multi-ethnic meta-analysis OR (95%CI) P value
Novel T2D locus													
rs9552911	13 23864657	*SGCG*	A/G	0.08	0	0.61 (0.47-0.80) p=3.08x10^{-4}	0.69 (0.59-0.82) p=1.14x10^{-5}	1.04 (1.00-1.07) p=0.037	1.00 (0.98-1.03) p=0.77	-	**0.67 (0.58-0.77) p=1.82x10^{-8}**	1.01 (0.98-1.05) p=0.47	1.01 (0.99-1.03) p=0.49
Previously reported T2D loci													
rs1470579	3 185511687	*IGF2BP2*	A/C	0.59	0.5	0.76 (0.66-0.88) p=2.53x10^{-4}	0.87 (0.8-0.95) p=1.87x10^{-3}	0.95 (0.93-0.97) p=1.79x10^{-6}	0.94 (0.92-0.97) p=7.21x10^{-6}	0.88 (0.92-0.84) p=2.17X10^{-9}	0.84 (0.78-0.9) p=5.02x10^{-6}	**0.94 (0.92-0.96) p=4.19x10^{-9}**	**0.94 (0.92-0.95) p=1.81x10^{-19}**
rs7903146	10 114758349	*TCF7L2*	T/C	0.35	0.28	1.31 (1.13-1.52) p=3.23x10^{-4}	1.50 (1.36-1.65) p=7.83x10^{-17}	1.13 (1.10-1.16) p=6.12x10^{-25}	1.12 (1.06-1.19) p=1.06x10^{-4}	1.40 (1.34-1.46) p=2.21X10^{-51}	**1.44 (1.33-1.56) p=3.32x10^{-19}**	**1.15 (1.13-1.18) p=2.71x10^{-35}**	**1.19(1.16-1.21) p=8.53X10^{-75}**

[a]-effect allele frequency, [b]- Sikhs, [c]- Euro Caucasians , All *P* values are two sided